DOCUMENT RESUME

ED 450 347 CS 014 271

AUTHOR McEneaney, John E.
TITLE New Approaches to Data Collection and Analysis in Online Reading Studies.
PUB DATE 1999-12-03
NOTE 7p.; Paper presented at the Annual Meeting of the National Reading Conference (Orlando, FL, December 1-4, 1999). Small type in figures may not reproduce well.
AVAILABLE FROM Full text at: http://www.oakland.edu/~mceneane/nrc/conf99/index.html.
PUB TYPE Reports - Descriptive (141) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Data Analysis; *Data Collection; Elementary Secondary Education; Higher Education; *Hypermedia; *Reading Processes; *Reading Research; *Research Methodology

ABSTRACT

This paper describes new approaches to data collection and analysis utilizing technology-assisted methods that are now possible in online studies of reading. Two techniques are described in detail, one supporting a non-intrusive method for real-time data collection during online reading, and a second describing a new method for visualizing and assessing user navigation in hypertext. Results of applying the methods in a series of empirical studies will be described along with suggestions for other applications. Contains 2 tables and 2 figures of data. (Author/RS)

Reproductions supplied by EDRS are the best that can be made from the original document.

ED 450 347

# New approaches to data collection and analysis in online reading studies.

John E. McEneaney, Ph.D. - Oakland University
mceneane@oakland.edu

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
J.E. McEneaney
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)



A Paper Presented at the Annual Meeting of the National Reading Conference. December 3, 1999.
Orlando, FL
Paper available online: http://www.oakland.edu/~mceneane/nrc/conf99/index.html

---

***Abstract*** - This paper describes new approaches to data collection and analysis utilizing technology-assisted methods that are now possible in online studies of reading. Two techniques are described in detail, one supporting a non-intrusive method for real-time data collection during online reading, and a second describing a new method for visualizing and assessing user navigation in hypertext. Results of applying the methods in a series of empirical studies will be described along with suggestions for other applications.

---

CS 014 271

BEST COPY AVAILABLE

**Introduction**

Although there has been growing appreciation in the reading research community that online technologies are altering the nature of reading and the roles of readers, there has been only limited consideration of ways new technologies of reading are altering the materials, methods, and concepts employed in reading research. That is the purpose of this paper - to focus on the ways new reading technologies are contributing to reading research, particularly in online environments. Specifically, the presentation will describe experimental materials, procedures, and analytic techniques that have arisen because of new tools and perspectives created with the development of online reading environments.

The presentation will focus on two technology-based methods that have considerable potential in online reading research and have been successfully implemented in a series of empirical studies. One method involves real-time data collection during online reading utilizing scripts that run as background programs while readers access reading materials with web browsers. A second set of methods provides a means for analyzing the real-time navigational data that is collected using the first set of methods. After describing the methods, the paper will describe how they were implemented in a series of empirical studies and what the results of those studies suggest about the application of these kinds of technologies in reading research, particularly in studies examining web-based materials.

**Real-time Data Collection**

Real-time data collection (RDC) has a substantial history in reading research in the forms of read-aloud protocols, records of eye-movements, and brain-scan studies, but online reading environments significantly enhance the potential of RDC to contribute to investigations of reading. One reason the significance of RDC is heightened in online environments is that these environments almost invariably rely on computers to deliver reading material and the prodigious computing power of modern technology means there are enormous "reserves" to support a wide range of other activities in addition to the relatively minor demands imposed by text management. In an age of inexpensive but high-powered multi-tasking computers, a machine that delivers reading materials has plenty of processing power remaining to gather, sort, store, and even analyze data in real time.

Another reason to expect that real-time data collection and analysis has potential for our immediate and long-term study of reading is the fact that more and more readers are becoming familiar with the specialized needs and demands of online environments and opportunities for data collection are increasing explosively as more and more readers venture out onto the web. Moreover, with the gradual emergence of (relatively) stable web standards for stand-alone (Java) and scripting languages (Javascript, JScript, VBScript) the ease with which programs can be developed to support real-time data collection are greatly enhanced. And development of background code is further simplified by a wide range of built-in browser functions that can be accessed controlled using by and standardized the browser Javascript engine

In a web-based environment the most difficult problem related to RDC has to do with recording data, usually as a result of web browser designs that intentionally prevent modification of hard disk files in order to avoid "infection" by web-borne computer viruses. There is, however, one significant exception to the general rule that "Browsers shall not write." The exception is known as the browser "cookie file", and while generally isolated from other files on a computer, this file can, in fact, be modified, and thus serves as a convenient and secure repository for data collected in online sessions, which is precisely how most commercial web sites use it.

An obvious choice, therefore, for a data file to record web-based activity is the cookie file and, fortunately, modifying this file is straightforward. This is especially true when the data of interest involves some kind of repeated measure of the same data "type" since, in this case, a session-long string of data values can be recorded to a single "cookie". One example that is especially relevant in the context of web materials is a record of pages accessed, a data set that essentially consists of a sequence of pages visited in a browsing session. Another example of an "associated" variety of data set is a sequence of time durations that represent the amount of time spent on each page represented in the pages visited data set. These were, in fact, two kinds of data recorded in a series of studies exploring web-based reading (McEneaney, 1998; in press) with an example data file (Table 1) illustrating "path" (pages visited) and "time" (duration of page visits) cookies written by the Javascript function presented in Table 2.

```
password 123456789
tLast 919219463220
pLast 6
path 6,13,14,78,14,6,19,6,25,6
time 5210,3570,2040,2740,990,...
```

Table 1: Cookie data recorded by the script in Table 2.

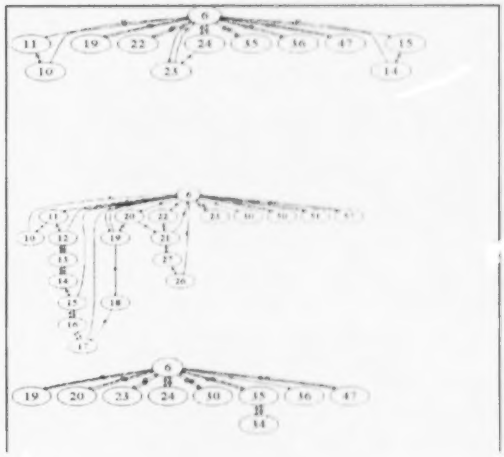

Figure 2. Less successful hypertext reader paths.

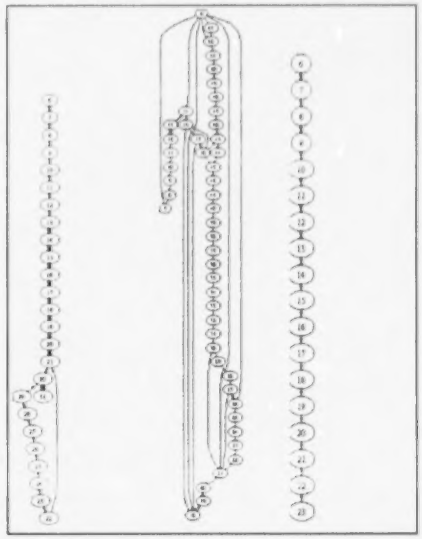

Figure 1. Successful hypertext reader paths

Researchers familiar with event-driven languages like HyperTalk associated with the HyperCard programming environment will note the similarity of the script presented in Table 2 but, even more importantly, Javascript code like that displayed in Table 2 can be built into a set of template pages that will allow even non-programmers to conduct RDC studies with little more technical expertise than that required to create web pages.

**Analysis of real-time data**

The data collected in online sessions does not always fit neatly into traditional analytic frameworks, particularly those that involve complex and sometimes circular sequences of data points, like those illustrated in Table 1. The usual solution has been to reduce the data sequence to a frequency count (Horney & Anderson-Inman, 1994; Schroeder & Grabowski, 1995). Unfortunately, this approach to analysis sacrifices a great deal of what is most important in the data by eliminating transitions between pages that capture user movement. In an effort to more adequately capture user movement, McEneaney (in press) has proposed methods that offer intuitively interpretable graphics and associated numerical measures that are empirically related to success in a hypertext reading task.

Specifically, results from a study examining the relationship of user navigation to success in a hypertext reading task associated distinctive graphic patterns of movement with successful and unsuccessful hypertext reading. Moreover these distinctive visual patterns were characteristic for both individuals in high and low achieving hypertext reading groups and held up when graphics were generated using data collapsed across subjects within groups. Examples of the resulting individual graphics are displayed in Figures 1 and 2.

Surprisingly, numerical measures reflecting the connectedness and linearity of these graphics (based on measures originally proposed by Botafogo, Rivlin, & Shneiderman, 1992) correlated significantly with success in the hypertext reading task although other measures of print reading ability and computing experience did not. This connection between user navigation and success in a hypertext reading task suggests that path data may be an important "window" on reading in online environments that could have implications both for research design in studies of online reading and hypertext design.

**Conclusions**

Although reading researchers seem to be aware that online reading requires new approaches to thinking about reading and literacy, these new ways of thinking have not often been generalized to include the ways we think about designing and conducting reading research, particularly in online environments. The evidence is mounting, however, that new technologies will transform reading research in ways that are every bit as dramatic as we have seen in reading practice.

The purpose of this session will be to familiarize NRC members with two ways new technologies are likely to influence reading research as ever larger numbers of readers turn to electronic media. Topics addressed in the session include real-time data collection and analytic techniques useful in displaying and characterizing user navigation in hypertext. After describing the general principles behind these topics, the presenter will report on a number of studies that employed these methods and indicate how these methods might be used in other ways. The session will conclude with an invitation to all participants to visit the online session web site associated with the roundtable presentation through links on the NRC web site (www.iusb.edu/~edud/EleEd/nrc/conf99). This online web site will provide participants an opportunity for further exploration of the methods described in the Orlando session and also allow participants to download software described in the presentation.

**References**

Botafogo, R. A., Rivlin, E., & Shneiderman, B. (1992). Structural analysis of hypertexts: Identifying hierarchies and useful metrics. ACM Transactions on Information Systems, 10(2), 142-180.

Horney, M. A. & Anderson-Inman, L. (1994). The ElectroText Project: Hypertext Reading Patterns of Middle School Students Journal of Educational Multimedia and Hypermedia, 3(1), 71-91.

McEneaney, J. (in press). Visualizing and assessing navigation in hypertext. Proceedings of the 10th Hypertext Conference of the Association for Computing Machinery. Darmstadt, Germany, February 21-25, 1999.

McEneaney, J. (1998). Are less able readers disadvantaged by reading in electronic environments? Proceedings of the 1998 International Symposium on Technology and Society, 28-32.

Schroeder, E. E., & Grabowski, B. L. (1995). Patterns of exploration and learning with hypermedia. Journal of Educational Computing Research, 13(4), 313-335.

```
function setCookie2(page) {
  var t = new Date();
  var t0 = t.getTime();
  var pLast = getCookie("pLast");
  var tLast = getCookie("tLast");
  var pathList = getCookie("path");
  var timeList = getCookie("time");
  var duration;
  if (tLast == null) {duration = 0} else {duration = t0 - tLast};
  var pathEntry;
  if (pathList == null) {pathEntry = page} else {pathEntry = ""+pathList+","+page};
  var timeEntry;
  if (timeList == null) {timeEntry = duration} else {timeEntry = "" +timeList+","+duration;};
  var oneWeek = 7*24*60*60*1000;
  var expDate = new Date();
  expDate.setTime(expDate.getTime()+oneWeek);
  if
(parseInt(page)!=78){document.cookie="pLast="+page+";expires="+expDate.toGMTString();}else{true
}
  document.cookie = "tLast="+t0+";expires="+expDate.toGMTString();
  document.cookie = "path="+pathEntry+";expires="+expDate.toGMTString();
  document.cookie ="time="+timeEntry+";expires="+expDate.toGMTString();
}
```

Table 2. JavaScript function that sets browser "cookies."